Optimal policy derivation for Transmission Duty-Cycle constrained LPWAN

Ruben M. Sandoval, *Student Member, IEEE,* Antonio-Javier Garcia-Sanchez, Joan Garcia-Haro, *Member, IEEE* and Thomas M. Chen, *Senior Member, IEEE,*

***Abstract*—Low-Power Wide-Area Network (LPWAN) technologies enable IoT devices to efficiently and robustly communicate over long distances, thus making them especially suited for industrial environments. However, the stringent regulations on the usage of certain ISM (Industrial, Scientific and Medical) bands in many countries in which LPWAN operate limit the amount of time IoT motes can occupy the shared bands. This is particularly challenging in industrial scenarios, where not being able to report some detected events might result in the failure of critical assets. To alleviate this, and by mathematically modeling LPWAN-based IoT motes, we have derived optimal transmission policies that maximize the number of reported events (prioritized by their importance) while still complying with current regulations. The proposed solution has been customized for two widely known LPWAN technologies: LoRa and Sigfox. Analytical results reveal that our solution is feasible and performs remarkably close to the Theoretical Limit for a wide range of network activity patterns.**

***Index Terms*—Long-Range Networks, LPWAN, IoT, LoRa, Sigfox, Markov Decision Process**

I. Introduction

THE Internet of Things (IoT) is now extensively employed in a wide variety of industrial environments, ranging from agriculture to power grids to manufacturing. The integration of sensing and actuating technologies helps increase productivity and efficiency, while at the same time, boosting control over the entire production chain. The increasing sophistication of IoT devices now enables robust and long-lasting networks with reduced installation and maintenance costs, which cut back on operational and capital expenditures.

From a technological point of view, industrial IoT deployments have taken over classical Wireless Sensor Networks (WSN) by implementing either low-power multi-hop communication standards (such as IEEE 802.15.4) or opting for cellular communication technologies (such as GSM/3G/4G). The former work in the Industrial, Scientific, and Medical radio bands (ISM), and are therefore license-free. However, due to their limited range, they tend to incur in high initial capital expenditures, since in order to cover medium to long distances, multi-hop topologies consisting of dozens of motes are required. This is even more so in the case of many industrial environments where the node density is not particularly high and the areas to be covered tend to be large, making most of these deployed motes serve merely as relying nodes in a multi-hop scheme [1]. On the contrary, IoT networks based on cellular approaches operate in licensed bands which unavoidably leads to higher operational expenditures, and force network operators to rely on Telcos.

However, in the past couple of years, Low-Power Wide Area Networks (LPWAN) have emerged as alternatives to the aforementioned options, especially for those environments in need of long-range communications, which is the case for many industrial scenarios like power and water grids, agricultural/rural areas or even cities. This type of network is characterized by three main features: (i) long coverage range (up to 30km [2]), (ii) low-power consumption (less than 100mJ per transmission [3]), and (iii) very long-standing deployment (motes have an expected lifespan of around 2 years with two AA batteries [3]). LPWAN are deployed following a classic star topology, thus removing the complexity and costs of a multi-hop infrastructure [1]. Furthermore, since they operate in ISM bands (normally low-frequency bands such as 868/902MHz for Europe/U.S.A.) and offer long-lasting features, maintenance costs are usually much lower compared to cellular alternatives.

Unfortunately, and as a consequence of operating in ISM bands, LPWAN networks are subject to strict legal regulations in many countries. One of the most inconvenient restrictions is the limit established on radio activity device patterns. For instance, in Europe, the ETSI EN 300 220-1 document [4] rules that the Transmission Duty Cycle (TDC) of devices that do not implement Listen-Before Talk mechanisms, as is the case of LPWAN networks, must fall below a certain value (normally expressed as a percentage, e.g. 1%)[1]. The Duty Cycle (DC) is defined as the percentage of time that a given node occupies a particular frequency band (measured over the length of an hour). Hence, a maximum Transmission DC of 1% implies that LPWAN motes have a maximum permitted TDC (Transmission Duty Cycle) of 36 seconds per hour. For example, motes cannot use a specific transmission band more than 36 seconds an hour, which corresponds to 1% of an hour. This TDC limitation also applies to other geographical zones and bands (e.g. the 779-787MHz band in China [1], or the 950-956MHz in Japan [5]) and, although it helps in reducing packet collisions, it potentially jeopardizes the ability of a network to effectively control valuable assets.

It is easy to think of situations in which nodes reporting non-important events (e.g. vibrations detected in a certain asset)

This research has been supported by the project AIM, ref. TEC2016-76465-C2-1-R (AEI/FEDER, UE). Ruben M. Sandoval also thanks the Spanish MECD for an FPU ref. FPU14/03424. Ruben M. Sandoval, Antonio-Javier Garcia-Sanchez, and Joan Garcia-Haro are with the Department of Tecnologías de la Información y Comunicaciones, Universidad Politécnica de Cartagena (UPCT), Antiguo Hospital de Marina, Cartagena Murcia 30202, Spain, e-mail: {ruben.martinez, antoniojavier.garcia, joang.haro}@upct.es. Thomas M. Chen is with School of Engineering and Mathematical Sciences, City University of London, EC1V 0HB, United Kingdom, email: tom.chen.1@city.ac.uk.

[1]Note that throughout the rest of the document, TDC will be used to denote the precise amount (in seconds) of time that nodes can access the medium per hour (e.g. 36 seconds), whereas DC will represent the percentage of such an hour (e.g. 1%).

may result in depleting the available TDC and thus rendering the mote unable to later give information about critical events (e.g. the global malfunction of such an asset). Therefore, it is crucial to recognize the TDC as one of the main performance-limiting factors in LPWAN-based deployments, and to manage the consumed transmission time as the scarce resource it is (as has been done in many works with the energy left in motes batteries [6]–[8] Unfortunately, and in contrast to battery management, the academic community still lacks exhaustive analyses that could result in optimal transmission policies in terms of the remaining TDC. Optimal transmission policies would be those that, taking into account the full context of the mote, would determine the action that maximizes the number of total high-importance packets transmitted over the entire node lifespan. The context of a mote could be described by the importance of the event being reported (e.g. low or high), the remaining TDC, propagation conditions, etc. Furthermore, and regarding the derived optimal action, some of the most widespread current LPWAN technologies, such as LoRa [9], allow the use of different configurations when transmitting packets. These configurations can increase the probability of successfully transmitting a packet at the cost of increasing the time-on-air of packets and hence, the consumed TDC. Therefore, an optimal transmission policy would also consider this trade-off and determine, in case of reporting an event, under which configuration it should be transmitted. It must be noted that, although the TDC-limitation resembles bandwidth restrictions, this is a very different problem. For instance, in the former, present decisions (e.g. transmit a packet) greatly influence future actions (e.g. not being able to transmit further packets) and the applied restrictions operate in time, independently of the number of raw bits transmitted. Conversely, under a bandwidth limitation, one could operate greedily, taking the action that maximizes the current figure of merit by, for example, limiting the amount of transmitted data and/or the speed at which such data is sent.

In the context of deriving action policies, and especially for WSN/IoT networks, the Markov Decision Process (MDP) framework stands out for its particularly well-suited form, optimal results and mathematical robustness [10]. The solution of MDP models are known as policies (in this case, transmission policies) and are guaranteed to indicate the best action to take at each moment according to the model of the environment. In the specific domain of decision making frameworks (such as MDP), the best action is conceptually defined as the optimal response in terms of of maximizing the total reward over the entire lifespan of a node. For the problem illustrated here, the total reward of a node is understood as the number of successfully transmitted events prioritized by their respective importance.

Although there are many works in the literature that apply the theory of decision making (and, in particular, MDP) to WSN/IoT networks [11]–[13], to the best of authors' knowledge, none of them has proposed any optimal transmission policy in TDC-limited networks. Hence, with the aim of filling the research gaps, we present the main contribution of this work: the derivation of an MDP-based transmission policy for TDC-constrained networks that:

- Complies with the TDC regulations of each geographical zone (e.g. max DC of 1%) and maximizes the number of successfully reported events.
- Yields the optimal action to be taken when a node is presented with the opportunity of reporting an event. Note that the set of actions might not be limited to transmitting or discarding the information, since several transmission configurations may be possible.
- Considers the importance of the packet when deriving the optimal action. Thus, high-importance packets should be prioritized over low-importance packets when deciding whether to send them or not, and under which available configuration.

The rest of the paper is organized as follows: The related work is presented in Section II. The analytical MDP formulation is first introduced in Section III, where a model of the problem is also characterized. In Section IV, the generic mathematical model is applied to the two currently most popular LPWAN technologies: LoRa [9] and Sigfox [14]. Section V validates the proposed models by comparing the average reward of a simulated industrial IoT network when: (i) our proposed MDP-based approach is used, and (ii) standard transmission policies are employed. After that, the results obtained from both approaches are compared to the maximum attainable rewards (i.e. the *Theoretical Limit* of the network) to further highlight the contributions. Finally, Section VI presents the conclusions and outlines future lines of research.

II. Related Work

Long range technologies such as LoRa [9] and Sigfox [14] have started to draw significant attention from the academic and industrial communities. Some of the published works in this field devote their efforts to analyzing the performance of real LPWAN deployments under different conditions: IoT devices monitoring civil infrastructures such as bridges [15], LoRa-based video surveillance systems [16], health monitoring motes [17], etc. On the other hand, some other studies are focused on analyzing the advantages, disadvantages, capabilities, and limits of the current implementations of these technologies from a technological point of view. For example, the real scalability of current LoRa networks [18], [19], the performance of their different configurations [20], and how these types of networks tolerate download traffic [21], amongst other things are being studied. Although they are very practical and illustrating, none of these works optimizes or analyzes the performance of LPWAN in a generic and theoretic fashion, which would allow their extrapolation to different technologies (LoRa, Sigfox, etc.) or their future implementations, beyond current transceivers. As a notable exception, [22] studied the impact of sub-band selection on LoRa motes by modeling nodes as an infinite, jockeying M/M/c queue (i.e. c servers, arrivals determined by a Poisson process, and exponentially distributed job services). Although the work is very well detailed, mathematically neat and applicable to future deployments, it does not capture the true, complex nature of real Long-Range networks, where resources are very scarce (i.e. infinite queues are impossible to implement) and traffic cannot always be assumed to follow a certain distribution.

Regarding the TDC-limitation problem, two works [1], [16] have recently highlighted the importance of TDC-aware networks by illustrating the problem of transmitting real-time video in Long-Range deployments. Although practical, the solution proposed focuses on deliberately breaking the 36s/hour

TDC limitation by complying with it in a network-aggregated fashion (i.e. the average network TDC is kept below 36s/hour, not the per-node TDC). In fact, [23] highlighted that the effects of TDC limitations jeopardize the actual capacity of large-scale deployments, and the only de-facto proposal to manage it, a fixed limit on the number of permitted messages per day, fails to provide the network with enough flexibility.

With the interest of contributing to fill the notable gap in research, we propose an approach to derive MDP-based transmission policies that fully comply with the TDC regulations while maximizing the number of high-priority reported events.

III. Mathematical Model

As commented on Section I, the goal of the system is to maximize the expected number of reported events prioritized by their importance, which is defined as the reward. To this end, we have opted for modeling the optimal action-derivation problem as an infinite horizon, discounted reward MDP [24]. First, because of the special ability of this mathematical framework to model the problem as an optimization task over discrete actions [25], which makes it particularly suitable for deriving transmission policies as explored by [8]. And secondly, because of the remarkable efficiency of current MDP-solving methods, which allows them to be used in very resource-constrained WSN/IoT motes, as [26] and [7] highlighted. The "infinite horizon" property indicates that the optimal policy is derived by considering the entire lifespan of motes. It is usually accepted that this otherwise infinite lifespan of motes may come to an end with certain small probability [6], [10]. This factors in the probability of node breakdown or manual disconnection, and globally represents the idea of maximizing almost immediate rewards instead of distant rewards, i.e. the "discounted reward" attribute of the MDP. For example, a reward of R obtained in the next hour is preferred to a reward of R obtained next year.

The proposed Markovian system operates in a discrete fashion (i.e. discrete time) in which time is subdivided into time slots with a length of Q seconds. Hence, the j-th time slot represents the time within seconds, with $j \in \mathbb{Z}$.

The working cycle of IoT/WSN networks deployed in industrial environments is usually ruled by the awake-asleep mote cycles. These motes usually remain in a dormant state to save energy, then wake up, sense some parameters and report them to a gateway, if needed, before going back to sleep. The periodicity with which nodes wake up and enter into the dormant state (i.e. the length of the awake-asleep cycle) equals T seconds or, similarly, T/Q time slots. Depending on how critical the controlled assets are, T should vary accordingly to allow for finer monitoring. Therefore, at the beginning of each k-th awake-asleep cycle, there might be a requirement to transmit an event e, i.e. sensor readings, (see Fig. 1). Each event of the arbitrary type i is generated with a probability λ_i. The set of the probabilities of generating any event conforms vector $\mathbf{\Lambda} = (\lambda_0, \lambda_1, ..., \lambda_P)$, with λ_0 being the probability of not generating any event and equal to $\lambda_0 = 1 - \sum_{i=1}^{i=P} \lambda_i$. Note that $0 \leq \lambda_i \leq 1$ and $\sum_{i=0}^{i=P} \lambda_i = 1$. Moreover, each packet reports an event of a different importance, which is represented by the priority of such an event $\mathbf{G} = (0, G_1, G_2, ..., G_P)$. The first item in the vector indicates the priority of reporting a non-generated event.

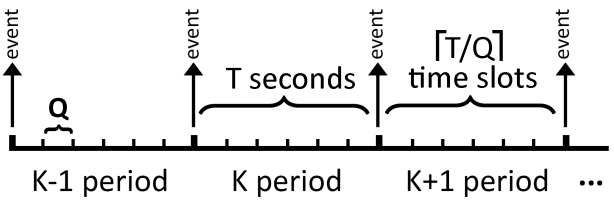


Fig. 1: Timing scale. awake-asleep cycles of T seconds are divided into T/Q time slots of length Q seconds each. Events can be generated at the beginning of each cycle, when sensor readings are obtained.

As introduced in Section I, over any given period of 1 hour, a node can occupy the transmission band up to $3600 * DC$ seconds or, equivalently, $\frac{3600*DC}{Q}$ time slots. The specific TDC consumption is determined by the transmission action, as some configurations lead to greater over-the-air times. Hence, when a given action a_i is taken, $C(a_i)$ time slots of TDC are expended. Considering up to $N+1$ different actions, the vector $\mathbf{C}$ models the time slot consumption of each action $\mathbf{C} = (0, C(a_1), C(a_2), ..., C(a_N))$, with the first action being not reporting the given event (or not having anything to report). Continuing with the dynamics of the TDC, at the beginning of each awake-asleep cycle, $Q_r = \frac{T*DC}{Q}$ time slots of TDC are obtained. This is derived from the fact that some time (T seconds) has passed and accordingly, some TDC has been "recharged". Note, that since it is particularly illustrating, we employ a terminology similar to the one found in energy consumption modeling. This defines a scheme in which the TDC is regarded as a commodity that can be consumed with transmissions and is recharged/regained (and stored) at the beginning of each cycle. Note that the maximum storable amount of TDC is equal to $Q_{MAX} = \frac{3600*DC}{Q}$. Some similarities can be observed between the proposed scheme and the classic Token Bucket algorithm [6]. A token is added to the bucket every $1/Q_r$ seconds, the bucket can hold at the most Q_{MAX} tokens, and events can be sent if there are enough tokens in the bucket to process their virtual length (defined by the vector $\mathbf{C}$). However, it is "time slots" and not packets (or bytes of packets) which are being stored. In fact, under the proposed system, packets will never be buffered or stored. Thus, they will be discarded if, at the time of being generated, they are not sent (either because there are not enough tokens/time slots or it is not in the interest of the future rewards).

As indicated in Section I, the generation and consumption of TDC work in a per-band fashion. That is, when different bands are available to motes, they must choose in which band each event will be transmitted. It is in such a band where the TDC will be consumed. However, since gateways (which are the core of the LPWANs) cannot be actively listening to different bands simultaneously, we will regard different bands as different networks. Hence, the model here introduced, which represents a specific network, shall be simply replicated if motes can operate in different networks (bands).

Then, the state s of any given mote can be fully described by the tuple (q_k, G_k) where q_k indicates the remaining TDC, in terms of time slots, at the beginning of the awake-asleep cycle k, and G_k denotes the priority of the event generated in the k-th cycle (including 0 if no event has been generated). Note that this process exhibits the Markov property, i.e., the

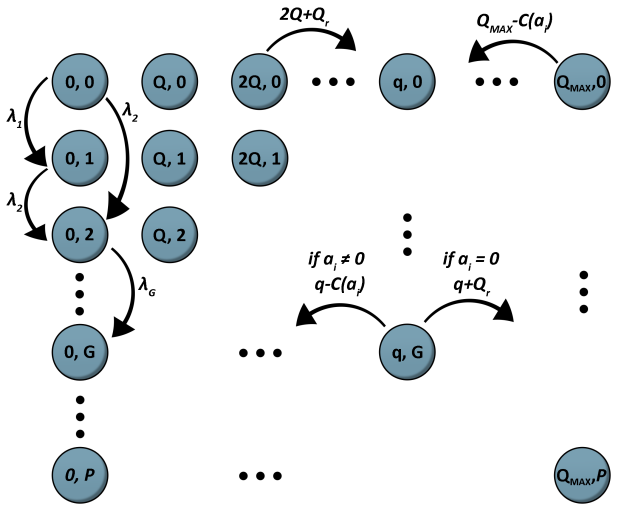


Fig. 2: State space of the proposed system. Transitions between states are depicted with arrows. Horizontal and vertical transitions are depicted independently for the sake of clarity.

state of the system is fully described by the current state and no past information is needed.

We can now represent the state space of the system as a two-dimensional Markovian model (see Fig. 2). The first dimension (horizontal axis) represents the accumulated TDC, whereas the second dimension (vertical axis) exemplifies the generation of an event. Horizontally, we move a deterministic number of time slots to the right every cycle: Q_r (the TDC recharge rate). Similarly, we move to the left when we transmit an event. The number of time slots we move in that direction depends on the action (specifically, on the TDC consumption of such an action defined by $\mathbf{C}$). Vertically, we move randomly following the event generation probability. At the beginning of each cycle, we move to a new event-generation position based on $\mathbf{\Lambda}$. Figure 2 illustrates this.

At any given state s, a certain set of actions, $\mathbf{A}$, are at the disposal of motes based on their remaining TDC, q_k, and the priority of the generated event G_k (that is, the state of such motes). Motes will act according to a transmission policy (defined by π) which should be regarded as a simple mapping between states and actions ($\pi : \mathbf{S} \rightarrow \mathbf{A}$). This policy determines what is called the value function $V(s, \pi)$, formally described as the expected total reward obtained when such a policy is followed given a starting state s:

$$V(s,\pi) = \mathbb{E}_s^{\pi}\left(\sum_{k=0}^{\inf} \gamma^k \cdot r(s_k, a_k)\right), s \in \mathbb{S}, \tag{1}$$

where k is the awake-asleep cycle index, a_k and s_k are the action taken and the state at that cycle respectively. $1-\gamma$ is the i.i.d (independent and identically distributed) probability of a node terminating its lifespan, again, due to a breakdown or a simple disconnection. Note that $\gamma \in [0, 1)$. $\mathbb{E}_s^{\pi}$ represents the expectation of the total reward under such policy. On the other hand, $r(s_k, a_k)$ denotes the reward obtained when action a_k is performed being at state s_k. This reward has been modeled to reflect the final goal: to successfully report events to the gateway (especially high-importance events). To appropriately consider this "success rate" in reporting events, it should be noted that, as indicated in Section I, some transmission configurations increase the probability of reception. Hence, the expected reward should increase accordingly. Thus, the reward is formally defined as follows:

$$r(s_k, a_k) = G_k \cdot PRR(a_k), \tag{2}$$

that is, the reward obtained when reporting a generated event is defined as the product of the priority of such event and the packet reception rate (PRR) under the employed configuration (i.e. the action). The function $PRR(a_k)$ depends on the particular technology and specific examples for different technologies will be given in Section IV.

The aim of the MDP is then to find the optimal policy π^* that maximizes the value function given an initial state:

$$\pi^* = \arg\max_{\pi} V(s, \pi). \tag{3}$$

This optimal policy π^* effectively maximizes the expected number of reported events (prioritized by their respective importance) over the entire mote lifespan, while complying with the TDC regulations at the same time.

Since the states, transitions, actions, and rewards structures of the system can be fully described, we have opted for tackling the optimal policy-derivation problem via model-based approaches (for which these four elements are needed). On the other hand, Reinforcement Learning (RL) alternatives depend on feedback signals to derive such policies. These feedback signals are typically modeled via ACKs in wireless networks. However, LPWAN gateways (in charge of acknowledging packets) are also TDC-limited. This poses a very strict limit on the number of ACKs that may be sent and thus, renders RL alternatives unfeasible. Unfortunately, traditional model-based approaches, like policy iteration or value iteration [25], require tabular representations of the system and thus, can only be applied if the number of states and actions is small. Therefore, to enable the application of such algorithms, Section IV elaborates on a set of simplifications that can be applied to the proposed model of Fig. 2, without any loss of precision, in real Long-Range networks.

Modeling the generation probability

As indicated above, to solve a model-based decision-making problem, the transition matrix of the MDP (i.e. how states evolve when actions are performed) must be known. Although the horizontal transitions in the two-dimensional MDP model represented in Fig. 2 are deterministically defined by the actions taken and the TDC obtained in each cycle (Q_r), the vertical transitions depend on the $\mathbf{\Lambda}$ vector (the event-generation probabilities). For the generic scenario in which events are being generated based on the sensor readings, the event-generation probabilities are unknown. Note that if events are generated periodically, the sensing rate must simply be adjusted to balance the consumption-generation of TDC. However, these probabilities can be effectively estimated by looking at the generation history of each mote. By considering the number (and type) of generated events over a given number of sensing cycles, we can roughly estimate the probability of packet generation. It is worth remarking that we only need an estimation of how likely is that an event is generated, not an estimation of when it will be generated. Nevertheless, large imprecisions in the estimation of Λ may translate, under some

circumstances, in a slight degradation of the performance of the derived policy –see appendix for further analysis on this degradation–. Since, at the beginning of a mote's life, this knowledge (the history of generated events) might be very scarce (especially for very rare events) we may incorporate some prior knowledge of the environment in order to start off with appropriate accuracy. For example, other motes deployed in the same scenario might provide valuable information regarding the probability of generating an event. Considering that events are independent to each other, the occurrence of a given type of event i can be regarded as a realization of a Bernoulli process with probability λ_i, which models how likely that event is to occur. To estimate λ_i, we can compute how likely the history of events $(e_1, e_2, ..., e_n)$ is under each value of $\lambda_i \in [0, 1]$ and weight it by our prior knowledge. Then, we can derive the value of λ_i that maximizes (λ_i^*) that product and thus, define the vector $\mathbf{\Lambda}^*$. Formally:

$$P(\lambda_i \vee e_1, e_2, ..., e_n) \propto P(e_1, e_2, ..., e_n \vee \lambda_i) \cdot P(\lambda_i) \quad (4)$$

$$\lambda_i^* = \arg\max_{\lambda_i} P(e_1, e_2, ..., e_n \vee \lambda_i) \cdot P(\lambda_i) \quad (5)$$

$P(\lambda_i)$ characterizes our prior belief about the distribution of a specific event generation probability. For instance, it seems reasonable to consider that high-importance events would tend to occur less frequently than low-importance events in well-functioning industrial environments.

Additionally, if Λ is known to change over time, that is, there is a seasonality component, a moving average can be used to let our knowledge about the event-generation process change over time. Furthermore, if events are known to be correlated, we can let $\lambda_{i,j}$ represent the probability of generating an event of type i after having generated an event of type j. Then, to estimate $\lambda_{i,j}$, we would only consider this type of sequence when analyzing the event history.

IV. Particularization of the Model

To assess the performance and benefits of the proposed solution, the MDP-based system model has been applied to the two most popular technologies in the current LPWAN networks arena: Sigfox and LoRa. The generic modeling (depicted in Fig. 2) is particularized to the specifics of Sigfox and LoRa technologies in subsections 4.a and 4.b, respectively. Furthermore, it is shown how the inherent properties of both technologies reduce the complexity of the MDP models (i.e. the size of the state and action space) and thus, allow them to be solved directly on actual IoT/WSN motes. In the next section (V), these models are used to simulate Sigfox/LoRa motes under a wide range of conditions. The rewards obtained with our proposal are compared with those obtained under different habitual policies.

Throughout the following subsections, the European 868-868.6MHz ISM sub-band [4], with its 1% DC limitation, is used. However, results can be directly applied to other bands without any loss of generality.

A. Sigfox

Sigfox operates both as an LPWAN technology (with a proprietary communication solution) and as a service provider (with its own LPWAN network, installed in 17 countries) [2]. Technology-wise, it employs ultra-narrow band (UNB) signals along with BPSK modulation to attain communication distances longer than 50km. Like most LPWAN technologies, it works in sub-GHz ISM bands, and provides a rather modest bitrate of 100 bps [27].

Parameter	Formula	Value
Q	N/A	0.05 seconds
T	N/A	5 seconds
T_{TX}	N/A	6 seconds
DC	N/A	1%
C	T_{TX}/Q	120 time slots (note that the TDC consumption for not sending any packet is 0)
Q_r	$T \cdot \frac{DC}{Q}$	1 time slot
Q_{MAX}	$3600 \cdot \frac{DC}{Q}$	720 time slots

TABLE I: Variables of the Sigfox's model

For the evaluation of the transmission policies, an IoT/WSN network deployment in an industrial environment is assumed. In such a scenario, it is usual that following the awake-asleep mote cycle, events are generated based on the pre-processed sensor readings. These readings reflect the value of certain assets (e.g. industrial machinery, soil moisture, rotor vibration, etc.) and thus, are reasonably supposed to all be equal in length. Since Sigfox only allows payloads of up to 12 bytes, this payload length is considered in this and following sections. The entire packet length, once the headers have been added, is 26 bytes (208 bits) in length, which at 100 bps, takes 2 seconds to be sent. However, to increase the packet reception rate, Sigfox transmits, by default, three times the same packet (one after another), thus occupying the working band for 6 seconds ($T_{TX} = 6$).

Note that Sigfox provides a single transmission configuration. Hence, reducing the action space to either transmit or do not transmit (a_1 and a_0 respectively). The time is divided into time slots of $Q = 0.05$ seconds. Thus, each 6-second transmission consumes 120 time slots of TDC. The length of the awake-asleep cycle is chosen to be $T = 5$ seconds. This value is consistent with the sensing rate of IoT/WSN motes deployed in industrial scenarios that do not require real-time monitoring (which is not attainable with current LPWAN solutions [23]) and yields a recharge rate of 1 time slot ($Q_r = 1$) per cycle. Similarly, the 1% DC implies a maximum storable TDC of 36 seconds or $Q_{MAX} = 720$ time slots. Table I specifies the variables of the model along with the formulae when applicable:

Concerning the event generation, 2 types of events, based on their priorities, have been considered as an example: low-importance and high-importance events. Note that this does not entail any loss of generality and the models presented in this work can be easily generalized to any number of different priorities. Thus, we define the $\mathbf{\Lambda}$ vector as follows: $\mathbf{\Lambda} = (\lambda_0, \lambda_1, \lambda_2)$. This reflects a scenario in which low-importance events can be regarded as readings lying within acceptable values (i.e. controlled assets are operating normally), whereas high-importance events would model reports on anomalies in the controlled assets. As explained in Section III, the rewards obtained by motes depend on both the importance of the transmitted event (1 or 2 for the low- and high-importance events, respectively) and the PRR. Unfortunately, the exact curves for deriving the PRR from the signal-to-noise ratio (SNR) are not currently available for Sigfox motes. Therefore,

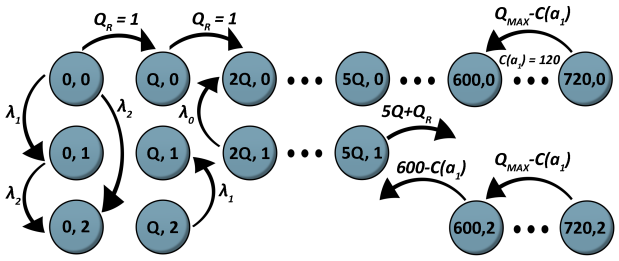


Fig. 3: State space of the Sigfox Markovian model. Transitions between states are depicted with arrows. Horizontal and vertical transitions are depicted independently for the sake of clarity.

the sensitivity of Sigfox gateways (-147dBm [28]) is rightfully employed to compute the PRR. That is, if the received power is over the sensitivity threshold, the PRR will be 1, and 0 otherwise.

The parameters in Table I configure the Sigfox system model depicted in Fig. 3. This system shows states with 721 different TDC values (from TDC=0 to TDC=Q_{MAX}=720) and 3 different G values (the type of the generated event). Hence, configuring a joint state-space of 2163 (721x3) different states. Conversely, and as indicated above, the action-space is limited to two different actions: report (a_1) and do not report (a_0) an event.

B. LoRa

LoRa is unarguably the main actor in the current LPWAN scene. It has the support of many worldwide technology leaders (Cisco, Microchip, IBM, HP, etc.) [29], and unlike Sigfox, it enables the deployment of private networks. From a technological point of view, LoRa offers a proprietary Chirp Spread Spectrum modulation (CSS) to achieve communication distances greater than 15km in sub-GHz bands [30]. Moreover, LoRa allows us to modify several transmission properties: the bandwidth and central frequency of the communication, the Coding Rate (CR), which is the ratio between the length of the packet and the length of the error-correction code, the Transmission Power, and the Spreading Factor (SF). This last parameter, the SF, is defined as the ratio between the symbol rate and chip rate. Higher spreading factors can increase the sensitivity and range of communication at the expense of increasing the over-the-air time of the packets. Thus, this parameter is fundamental in the derivation of the optimal transmission policy and leads to a trade-off which is worth studying. Higher SF factors increase the PRR and hence, the expected immediate reward, but consume more TDC, reducing the potential future rewards. Similarly, the Coding Rate also affects both the TDC consumption, as decreasing it increases the effective length of the packet, and the expected reward, since with higher values, the PRR increases as the communication gets more resistant to errors. The relation between the SF/CR configurations and the bit-error-rate and therefore, the PRR, has been modeled in [31] and is illustrated in the following expression:

$$BER = 10^{\alpha e^{\beta SNR}}, \tag{6}$$

CR	SF	a_i	α	β	TXR
4/5	7	a_1	-30.2580	0.2857	3410 bps
	8	a_2	-77.1002	0.2993	1841 bps
	9	a_3	-244.6424	0.3223	1015 bps
	10	a_4	-725.9556	0.3340	507 bps
	11	a_5	-2109.8064	0.3407	253 bps
	12	a_6	-4452.3653	0.3317	127 bps
4/7	7	a_7	-105.1966	0.3746	2663 bps
	8	a_8	-289.8133	0.3756	1466 bps
	9	a_9	-1114.3312	0.3969	816 bps
	10	a_{10}	-4285.4440	0.4116	408 bps
	11	a_{11}	-20771.6945	0.4332	204 bps
	12	a_{12}	-98658.1166	0.4485	102 bps

TABLE II: α, β, and TXR parameters for different values of CR and SF

Parameter	Formula	Value
Q	N/A	0.051 seconds
T	N/A	5 seconds
TXR	N/A	(3410, 1841, 1015, 507, 253, 127, 2663, 1466, 816, 408, 204, 102) bps
DC	N/A	1%
C	$C(a_i) = \frac{208/TXR_{a_i}}{Q}$	(1, 2, 4, 8, 16, 32, 2, 3, 5, 10, 20, 40) time slots (note that the TDC consumption for not sending any packet is 0)
Q_r	$T \cdot \frac{DC}{Q}$	5 time slots
Q_{MAX}	$3600 \cdot \frac{DC}{Q}$	706 time slots

TABLE III: Variables of the LoRa's model

where alpha and beta depend on the specific configuration of the SF and CR, covered in Table II. In turn, the PRR, used to model the rewards as per Eq. 2, can be computed as follows:

$$PRR = (1 - BER)^L, \tag{7}$$

where L is the length of the packet. As indicated above, the specific configuration of SF/CR also affects the transmission rate of packets. For a bandwidth of 125KHz, which is the most common configuration, the effective transmission rates (TXR) are included in Table II for all the 6 configurations of SF available, and the 2 different values of CR studied (the two CR configurations for which the α and β parameters were available).

If the time slot length (Q) is set to 0.051 seconds, and again, 26-byte packets are considered (208 bits), the time slot consumption of each action can be easily computed as follows (note that the SF/CR configurations define the action set):

$$C(a_i) = \frac{T_{TX}}{Q} = \frac{208/TXR_{a_i}}{Q} \tag{8}$$

Similarly, for an awake-asleep cycle of 5 seconds, the time slot recharge rate equals 1 time slot ($Q_r = 1$) and the maximum storable TDC 706 time slots ($Q_{MAX} = 706$). Table III summarizes the variables of the model along with the formulae when applicable.

Technically, LoRa only defines the physical layer of the wireless communication, whereas LoRaWAN standardizes the upper layers. Within these layers, the Medium Access Control (MAC) sublayer ensures compliance with TDC regulations in a very specific fashion. Instead of simply enforcing the maximum transmission duty cycle of, for example, 1% per hour (i.e. 36 seconds), it forces the motes to undergo a silent period after each packet transmission. Therefore, the TDC

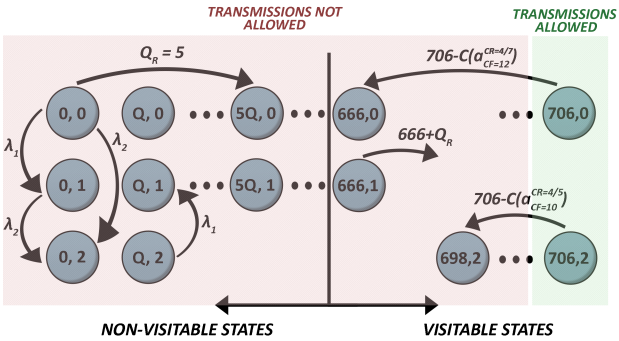


Fig. 4: State space of the LoRa Markovian model. Only when motes reach the end of the chain (at the right-hand side of the model) can new packets can be. Therefore, states with fewer than 666 time slots will never be visited.

regulation is not only met "per hour", but also between any two transmitted packets. This silent period (T_{off}) depends on how long the radio was ON to transmit the last packet (T_{TX}) and it is formally defined as:

$$T_{off} = \frac{T_{TX}}{DC} - T_{TX} \tag{9}$$

Although this helps in reducing the number of packet collisions, it has the undesirable effect of preventing LoRa motes from being able to transmit packets in bursts. Furthermore, this alters the way in which the model (Fig. 2) makes a transition. Instead of progressively "accumulating" TDC each cycle, a mote would lie at the end of the TDC chain (at the right-hand side of it, where Q_{MAX} is in Fig. 2) and every time it generates a packet, it moves the number of time slots defined by C to the left (see Fig. 4). Only then, would the mote start to progressively move to the end of the chain again (Q_r time slots each cycle). And when it finally reaches the end of the chain, another packet can be sent again (i.e. when the mote is not at the end of the chain, it is considered to be in its T_{off} period and hence, cannot send any further packets). This behavior results in the leftmost side of the chain never being visited. In particular, the states to the left of the value $Q_{MAX} - C_{MAX} = 666$ time slots (see Fig. 4) will not be reached. Note that $C_{MAX} = 40$ time slots and corresponds to $C(a_{12})$. Therefore, these states can be omitted to speed up the computation of the MDP solution.

Thus, the state-space of the LoRa model is simplified to only account for TDC values ranging from 666 to 706 (41 different values) and event generation from 0 to 2 (3 different values). Hence, the entire system model contains 123 different states (41*3) and 13 different actions (not sending a packet plus the 12 different combinations of SF and CR). Note that LoRa motes operating in the 868-868.6MHz sub-band can select one of the 3 different channels to transmit a packet. However, since they belong to the same sub-band, the TDC consumption is aggregated among them, and the system dynamics are not altered at all. Therefore, TDC-wise motes are assumed to select a channel randomly when facing a new transmission, which is the current by-default behavior of LoRaWAN motes as per the standard.

V. Performance analysis of the transmission policies

To analyze and judge the performance obtained when the proposed MDP-based transmission policy is implemented in LPWAN motes, two metrics are employed: (i) a measure of how well motes could possibly do (i.e. the maximum attainable reward, defined by Eq. 11), and (ii) a measure of how well motes do with other policies (the reward obtained by following other transmission policies, defined by Eq. 10). The former represents the highest performance a node can attain if it knew beforehand all the events that would be generated over a fixed time period of K sensing cycles. Mathematically, it represents a *Theoretical Limit* on the attainable rewards, and can be computed as the maximization of the sum of discounted rewards over the set of available actions for the K cycles (Eq. 11). Although knowing what events will be generated beforehand is not realistic, it offers us a clear insight into the *regret* of using a certain transmission policy. This figure of merit (the *regret*) is formally defined as the difference between the maximum attainable reward (R^{MAX}) and the reward obtained with a given transmission policy (R, –Eq. 10, which is a particularization of the Eq. 1 for a fixed time period and a given event history–).

$$R = \sum_{k=0}^{K} \gamma^k r(s_k, a_k), \text{ with } a_k = \pi(s_k) \tag{10}$$

$$R^{MAX} = \max_{a_k} R, \text{ subject to } TDC_{k+1} = min(Q_{MAX}, TDC_k + Q_r - C(a_k)) \geq 0 \tag{11}$$

Furthermore, in order to have a good idea of the true performance of the MDP-based transmission policy under different scenarios, our proposal is compared to two different intuitive policies:

- Always Transmit (AT): Every generated event is transmitted if enough TDC is available. This is the by-default policy in scenarios where the priorities of events are disregarded, as is the case in the majority of current LPWAN deployments. Every generated event is treated the same way and transmitted if possible.
- Transmit High-Importance Events Only (THIEO): Only high-importance events will be transmitted (if enough TDC is available). In very TDC-restricted networks (like networks with 1% of DC) it is tempting to save such TDC for the dispatching of high-importance events. This policy is aimed at guaranteeing the transmission of high-importance events by allocating all the available TDC to them.

Since the rewards obtained depend on the history of generated events, all the policies will be tested for different values of $\mathbf{\Lambda}$. Let $\theta = \frac{\lambda_1}{\lambda_2}$ be the relation between λ_1 and λ_2 (the probabilities with which low and high-importance events are generated respectively). By varying θ, different types of networks can be characterized. For instance, $\theta = 2$ represents a network in which low-importance events, events containing readings that lie within normal/expected values, are generated twice as often as high-importance events (reports on anomalies or unexpected values). Furthermore, let $\phi = \lambda_1 + \lambda_2 \in [0, 1]$ be the probability of which an event of any type is generated during every sensing cycle. This parameter adjusts the *activity*

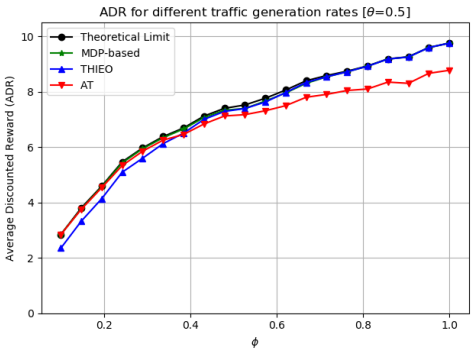
of the network. For instance, $\phi = 0.6$ indicates that with a 60% probability, an event is generated at any given cycle. These two parameters (ϕ and θ) define the activity patterns of the network and particularize λ_1 and λ_1 (e.g., for $\theta = 2$, $\phi = 0.6$, the following values are derived: $\lambda_1 = 0.4$ and $\lambda_2 = 0.2$).

For the evaluation of the different policies, ϕ has taken values from 0 to 1 in steps of 0.05 (20 different values that cover a wide range of networks between very quiet to very active ones). On the other hand, θ has been set to the following discrete values: 0.5, 1, 2, and 4 to analyze four different LPWAN networks according to the relation between both types of events. For each of these 80 combinations (20 different values of ϕ and 4 different values of θ), 1000 sensing/awake-asleep cycles of 5 seconds each have been simulated with 20 different seeds to compute the Average Discounted Reward, ADR, (the average of Eq. 10 over 20 different randomly generated event-histories). For the MDP-based transmission policy, the estimated generation rates ($\mathbf{\Lambda}^*$) are updated with past experiences every 50 sensing cycles, following the procedure described in Subsection III.B. It should be mentioned that the Python code employed to run the simulations has been made publicly available to support the research in this area [32].

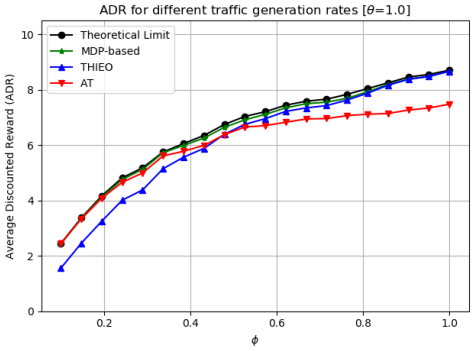
A. Sigfox

The difference in the ADR (i.e. *regret*) between the maximum attainable reward (i.e. the *Theoretical Limit*) and the proposed solution (denoted as MDP-based) is remarkably low for the entire space of ϕ and θ. In fact, the greatest difference found is 0.186 for ϕ=0.71 and θ=4 (this means ADR values of, in the worst case, just 3.12% smaller than the *Theoretical Limit*). However, these differences increase up to 1.23 for the AT policy and 2.24 for the THIEO policy (ϕ=1.0, θ=1 and ϕ=0.24, θ=4 respectively), yielding ADR values 14.09% and 58.96% smaller than the *Theoretical Limit* respectively. Such increments in the *regret*, reflect the fact that only with the MDP-based proposed transmission policy, can motes effectively adapt to the changing conditions of a realistic network, swiftly finding a suitable event-reporting policy that boosts their performance. Ultimately, this translates into a higher number of events being reported, always in a way in which their respective priorities are considered.

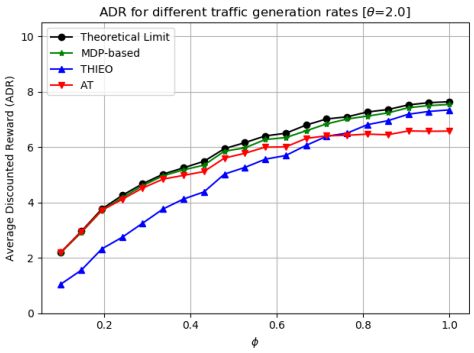
When motes generate very few events (that is, for low values of ϕ), the AT policy gets very close to the *Theoretical Limit*. In these situations, the TDC is accumulated faster than it is consumed, and thus, every event can (and should) be transmitted. This is precisely what AT policy does. On the other hand, when the network is very active (high values of ϕ), the rewards obtained under the THIEO policy (which only reports high-importance events) also tend to reach the *Theoretical Limit* (especially for θ=0.5, when more high-importance packets are generated). The intersection between AT and THIEO rewards occurs at different points of ϕ based on the exact value of θ. For instance, high values of θ entail a higher proportion of low-importance events and thus, make the THIEO policy tend to drop too many events by discarding low-importance ones. Moreover, and as a secondary effect, when more high-importance events are generated (i.e. when θ grows), the global ADR tends to increase. This is reflected by the fact that ADR values are generally greater for θ=0.5 than for θ=4.



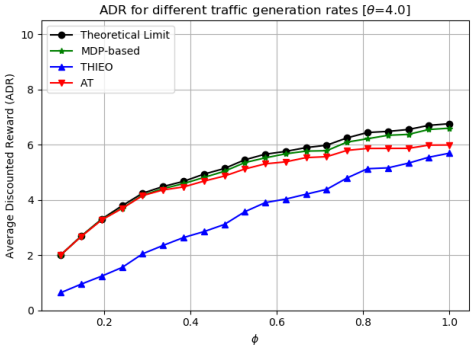
(a)



(b)



(c)



(d)

Fig. 5: ADR values obtained with Sigfox motes for (a) $\theta = 0.5$, (b) $\theta = 1.0$, (c) $\theta = 2.0$, and (d) $\theta = 4.0$

Nevertheless, for the Sigfox technology, neither AT nor THIEO policies can effectively adapt to the ever-changing conditions of industrial LPWAN networks, and thus, they would perform very poorly in dynamic scenarios, which are the most common type of industrial environment. Conversely, the proposed, MDP-based policy efficiently controls the behavior of LPWAN motes to adapt their transmission patterns to fluctuating conditions of the environment, attaining near-optimal performance for a wide range of situations.

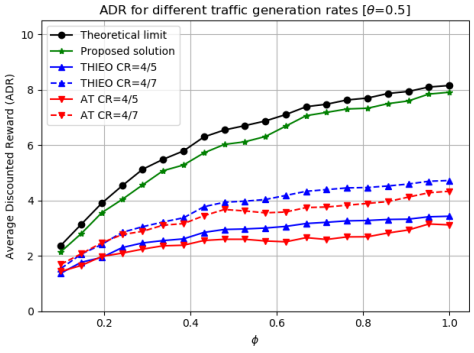
B. LoRa

Since in LoRa, packets can be sent under different configurations, for the AT and THIEO policies, the Spreading Factor is adjusted with the Adaptive Data Rate algorithm [33]. This algorithm is part of the LoRaWAN recommendations and it is basically in charge of finding the most suitable SF for the given conditions (specifically, the perceived SNR). With respect to the CR, the two configurations (4/5 and 4/7) have been exhaustively evaluated. On the other hand, the bandwidth and the transmission power, which do not affect the TDC consumption, have been set to their default values: 125kHz and 30dBm respectively.

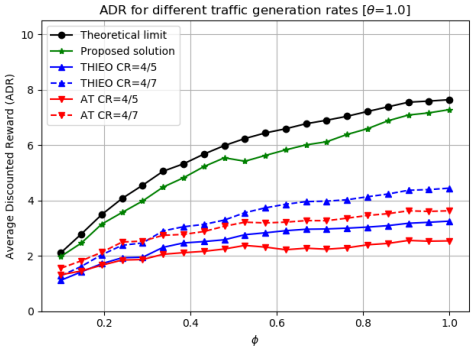
The first notable fact in Fig. 6 is the large difference between the ADR obtained under the MDP-based policy and the ADR achieved under the AT and THIEO intuitive policies. The maximum *regret* yielded by the AT policy is 5.1 for CR=4/5, and 4.01 for CR=4/7 (both for ϕ=θ=1.0). ADR values are 68.58% and 55.14% smaller than the *Theoretical Limit*, respectively. In turn, the maximum *regret* obtained under the THIEO policy is 4.7 for CR=4/5 and 3.43 for CR=4/7 (both for ϕ=1 and θ=0.5). ADR values are 60.05% and 50.07% smaller than the *Theoretical Limit*, respectively. These rather large differences reveal the importance of reckoning with a policy that adapts to environmental conditions, especially when the set of possible actions is fairly large. In LoRa, 13 different actions are considered. THIEO and AT policies, limited to employing a fixed set of two actions, either not transmitting or transmitting under a fixed configuration determined by the Adaptive Data Rate algorithm, cannot regulate mote behavior to meet the network conditions. As in the Sigfox case, THIEO tends to perform better than AT for high values of ϕ and especially for low values of θ. In other words, when either a large proportion of events are high-importance events, or a large number of them are generated in absolute terms, it is advisable to employ the TDC only for sending high-importance events.

In contrast, the proposed MDP-based solution seems to perform relatively well under any circumstances, with the maximum *regret* being 1.18 for ϕ=0.85 and θ=2. This means that ADR values are, in the worst case, 18.21% smaller than the *Theoretical Limit*. The ability of the proposed solution to choose the best SF and CR configurations in each scenario results in a higher capacity to adapt to different situations, hence, being able to effectively report a larger number of prioritized events.

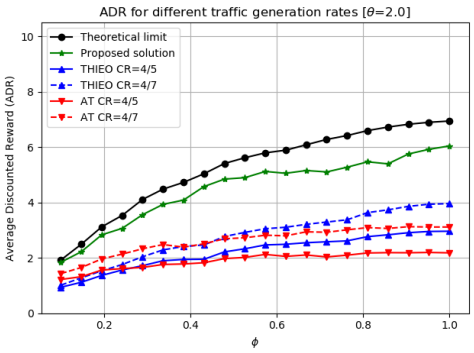
It is worth noting that the differences appreciated between Sigfox and LoRa (in terms of ADR), stem from the fact that LoRa enforces a silent period T_{off} after every transmission. This makes that a larger portion of packets get sent later in time. With a discount rate γ strictly smaller than 1 –the common practice–, postponing transmissions entails a



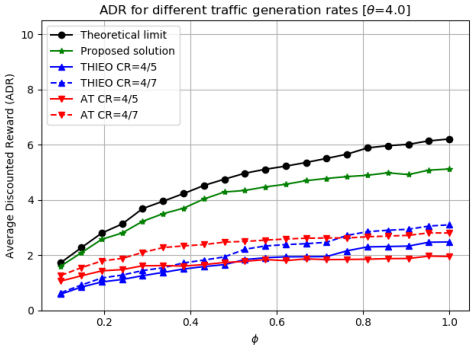
(a)



(b)



(c)



(d)

Fig. 6: ADR values obtained with LoRa motes for (a) $\theta = 0.5$, (b) $\theta = 1.0$, (c) $\theta = 2.0$, and (d) $\theta = 4.0$

penalization in the obtained reward (precisely, the reward is scaled by γ). In fact, it can be shown how, if γ approaches one, the obtained ADR values under LoRa surpasses those obtained under Sigfox –the average non-discounted rewards are generally larger for LoRa than for Sigfox–.

Regarding the complexity of the proposed transmission policy for both, LoRa and Sigfox motes, it should be noted that traditional MDP-solving methods, such as Value Iteration, have been formally proven to run in polynomial time with respect to the number of states and actions for a fixed discount rate γ [34]. Moreover, the memory requirement has been shown to be lineal with the number of states [35]. These two properties make the proposed MDP-based solution very well-suited for running in resource-constrained devices, such as IoT motes and in time-critical scenarios, such as industrial environments. Furthermore, the obtained transmission policies can be easily stored in 1D arrays in motes. These are extremely efficient and fast memory structures. In fact, for the worst-case scenario (LoRa), the complete policy could be stored in less than 62 bytes (123 different states multiplied by 4 bits per state).

VI. Conclusions

Long-Range IoT networks have started to draw the attention of the academic and industrial communities as very promising alternatives to other classic IoT technologies. This is mainly due to their longer communication ranges, robustness, simplicity, and convenient use of the license-free ISM bands. However, the limitations imposed on ISM bands in many countries hinder the ability of Long-Range IoT motes to make free use of the shared medium. In particular, the restriction to the amount of time motes can occupy the ISM bands (normally less than 36 seconds an hour) might jeopardize the ability of these kinds of networks to operate in industrial scenarios, where sensed data must flow in a timely fashion.

To alleviate this situation, an optimal transmission policy based on the analytical framework of Markov Decision Processes (MDP), has been derived with two objectives: (i) to maximize the number of reported events, prioritized by their importance, and (ii) to comply with the ISM regulations. This has been accomplished in two steps. Firstly, a general model of Long-Range IoT motes has been proposed and, secondly this model has been tuned for two widely-known technologies: LoRa and Sigfox. Motes of both technologies have been simulated under different network conditions. Unlike other traditional policies tested, whose performance is strongly coupled with network conditions, our obtained results reveal that the proposed solution performs very close to the maximum Theoretical Limit under almost any condition. Furthermore, the proposed solution is computationally fitted for resource-constrained motes and for time-critical scenarios, thus making it a good solution for IoT motes deployed in industrial scenarios.

Appendix
Impact of the estimation of Λ on the ADR

For the proposed MDP-based approach, the event-generation rate, Λ, need to be estimated. When the estimated Λ largely diverges from the true underlying event-generation rate, the derive transmission policies might be sub-optimal. We have analyzed this by purposely introducing perturbations in the estimated Λ to later analyze the obtained performance degradation. This degradation has been computed as a difference between the ADR obtained with the true Λ and the ADR obtained with the perturbed Λ. In turn, the perturbed Λ has been generated by adding zero-mean Gaussian noise to the true Λ.

Results show that, even in the presence of large perturbations (noise values that alter the estimated Λ in more than 50%), the degradation in the ADR varies between 0% to 18% depending on the specific scenario.

References

[1] C. Pham, "QoS for Long-Range Wireless Sensors Under Duty-Cycle Regulations with Shared Activity Time Usage," *ACM Trans. Sen. Netw.*, vol. 12, no. 4, pp. 33:1—-33:31, sep 2016. [Online]. Available: http://doi.acm.org/10.1145/2979678

[2] J. Petajajarvi, K. Mikhaylov, A. Roivainen, T. Hanninen, and M. Pettissalo, "On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology," in *2015 14th International Conference on ITS Telecommunications (ITST)*, dec 2015, pp. 55–59.

[3] M. Bor, J. Vidler, and U. Roedig, "LoRa for the Internet of Things," in *Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*, ser. EWSN '16. USA: Junction Publishing, 2016, pp. 361–366. [Online]. Available: http://dl.acm.org/citation.cfm?id=2893711.2893802

[4] ETSI, "Final draft ETSI EN 300 220-1 V2.4.1 (2012-01)," ETSI, Tech. Rep. REN/ERM-TG28-434, 2012. [Online]. Available: http://www.etsi.org/deliver/etsi_en/300200_300299/30022001/02.04.01_40/en_30022001v020401o.pdf

[5] S. Aust, R. V. Prasad, and I. G. M. M. Niemegeers, "IEEE 802.11ah: Advantages in standards and further challenges for sub 1 GHz Wi-Fi," in *2012 IEEE International Conference on Communications (ICC)*, jun 2012, pp. 6885–6889.

[6] W. Zhu, P. Xu, M. Zheng, G. Wu, and H. Wang, "Transmission Policies for Energy Harvesting Sensors Based on Markov Chain Energy Supply," in *Proceedings of the 10th EAI International Conference on Body Area Networks*, ser. BodyNets '15. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015, pp. 228–232. [Online]. Available: http://dx.doi.org/10.4108/eai.28-9-2015.2261406

[7] P. Blasco, D. Gunduz, and M. Dohler, "A Learning Theoretic Approach to Energy Harvesting Communication System Optimization," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1872–1882, apr 2013.

[8] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2988–3001, 2013.

[9] "LoRa," 2017. [Online]. Available: https://www.lora-alliance.org/

[10] M. A. Alsheikh, D. T. Hoang, D. Niyato, H. P. Tan, and S. Lin, "Markov Decision Processes With Applications in Wireless Sensor Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 3, pp. 1239–1267, 2015.

[11] Z. Ye, A. A. Abouzeid, and J. Ai, "Optimal Stochastic Policies for Distributed Data Aggregation in Wireless Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1494–1507, oct 2009.

[12] X. Fei, A. Boukerche, and R. Yu, "An efficient Markov decision process based mobile data gathering protocol for wireless sensor networks," in *2011 IEEE Wireless Communications and Networking Conference*, mar 2011, pp. 1032–1037.

[13] S.-T. Cheng and T.-Y. Chang, "An adaptive learning scheme for load balancing with zone partition in multi-sink wireless sensor network," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9427–9434, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741741200382X

[14] "Sigfox," 2017. [Online]. Available: http://www.sigfox.com

[15] F. Orfei, C. B. Mezzetti, and F. Cottone, "Vibrations powered LoRa sensor: An electromechanical energy harvester working on a real bridge," in *2016 IEEE SENSORS*, oct 2016, pp. 1–3.

[16] C. Pham, "Deploying a pool of long-range wireless image sensor with shared activity time," in *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, oct 2015, pp. 667–674.

[17] J. Petäjäjärvi, K. Mikhaylov, M. Hämäläinen, and J. Iinatti, "Evaluation of LoRa LPWAN technology for remote health and wellbeing monitoring," in *2016 10th International Symposium on Medical Information and Communication Technology (ISMICT)*, mar 2016, pp. 1–5.

[18] M. C. Bor, U. Roedig, T. Voigt, and J. M. Alonso, "Do LoRa Low-Power Wide-Area Networks Scale?" in *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '16. New York, NY, USA: ACM, 2016, pp. 59–67. [Online]. Available: http://doi.acm.org/10.1145/2988287.2989163
[19] O. Georgiou and U. Raza, "Low Power Wide Area Network Analysis: Can LoRa Scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, apr 2017.
[20] M. Bor and U. Roedig, "LoRa Transmission Parameter Selection," in *Proceedings of the 13th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), Ottawa, ON, Canada*, 2017, pp. 5–7.
[21] A. I. Pop, U. Raza, P. Kulkarni, and M. Sooriyabandara, "Does Bidirectional Traffic Do More Harm Than Good in LoRaWAN Based LPWA Networks?" *CoRR*, vol. abs/1704.0, pp. 1–6, 2017. [Online]. Available: http://arxiv.org/abs/1704.04174
[22] R. B. Sørensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of Latency and MAC-layer Performance for Class A LoRaWAN," *IEEE Wireless Communications Letters*, vol. PP, no. 99, p. 1, 2017.
[23] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the Limits of LoRaWAN," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017.
[24] M. Wiering and M. Van Otterlo, "Reinforcement learning," *Adaptation, Learning, and Optimization*, vol. 12, 2012.
[25] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, N. Roy, J. P. How, and Others, "A tutorial on linear function approximators for dynamic programming and reinforcement learning," *Foundations and Trends®in Machine Learning*, vol. 6, no. 4, pp. 375–451, 2013.
[26] K. Shah and M. Kumar, "Distributed Independent Reinforcement Learning (DIRL) Approach to Resource Management in Wireless Sensor Networks," in *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*, oct 2007, pp. 1–9.
[27] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A Study of LoRa: Long Range & Low Power Networks for the Internet of Things," *Sensors*, vol. 16, no. 9, 2016.
[28] W. Yang, M. Wang, J. Zhang, J. Zou, M. Hua, T. Xia, and X. You, "Narrowband Wireless Access for Low-Power Massive Internet of Things: A Bandwidth Perspective," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 138–145, 2017.
[29] LoRa-Alliance, "The Internet of Things - An explosion of Connected Possibility," LoRa, Tech. Rep., 2016. [Online]. Available: https://docs.wixstatic.com/ugd/eccc1a_de5fda268ed945e885a43a39b387528
[30] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.
[31] F. V. D. Abeele, J. Haxhibeqiri, I. Moerman, and J. Hoebeke, "Scalability analysis of large-scale LoRaWAN networks in ns-3," *CoRR*, vol. abs/1705.0, 2017. [Online]. Available: http://arxiv.org/abs/1705.05899
[32] R. M. Sandoval, A.-J. Garcia-Sanchez, J. Garcia-Haro, and T. M. Chen, "Python scripts." [Online]. Available: http://labit501.upct.es/ rmartinez/policies/
[33] LoRa-Alliance, "A technical overview of LoRa and LoRaWAN," Tech. Rep. [Online]. Available: https://www.tuv.com/media/corporate/products_1/electronic_components_and_lasers/TUeV_Rheinland_Overview_LoRa_and_LoRaWANtmp.pdf
[34] M. L. Littman, T. L. Dean, and L. P. Kaelbling, "On the Complexity of Solving Markov Decision Problems," *CoRR*, vol. abs/1302.4, 2013. [Online]. Available: http://arxiv.org/abs/1302.4971
[35] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, N. Roy, and J. P. How, "A Tutorial on Linear Function Approximators for Dynamic Programming and Reinforcement Learning," *Found. Trends Mach. Learn.*, vol. 6, no. 4, pp. 375–451, 2013. [Online]. Available: http://dx.doi.org/10.1561/2200000042

Ruben M. Sandoval (M'12) received the B.S. degree in telematics engineering in 2011 and M.S. degree in telecommunication engineering in 2014 from the Universidad Politecnica de Cartagena (UPCT), Cartagena, Spain. He is currently pursuing the Ph.D. degree at Universidad Politecnica de Cartagena (UPCT), Cartagena, Spain. His research interest includes efficient wireless sensor networks, IoT and the applicability of both network paradigms to manage critical environments.

Antonio-Javier Garcia-Sanchez received the M.S. degree and Ph.D. degree from the Universidad Politecnica de Cartagena, Spain in 2000 and 2005 respectively. Since 2001, he has joined the Department of Information Technologies and Communications (DTIC), UPCT. He is a (co)author of more than 50 conference and journal papers, fifteen of them indexed in the Journal Citation Report (JCR). He has been the main head in several research projects in the field of communication networks and optimization. His main research interests are wireless sensor networks (WSNs), and Smart Grids.

Joan Garcia-Haro (M'91) received the M.S and Ph.D degrees in telecommunication engineering from the Universitat Politecnica de Catalunya, Spain, in 1989 and 1995 respectively. He is currently a Professor with the Universidad Politecnica de Cartagena. He is author or co-author of more than 80 journal papers mainly in the fields of switching, wireless networking and performance evaluation. Dr. Garcia-Haro served as Editor-in-Chief for the IEEE Global Communications Newsletter, included in the IEEE Communications Magazine, from April 2002 to December 2004. He has been Technical Editor of the same magazine from March 2001 to December 2011. He also received an Honorable Mention for the IEEE Communications Society Best Tutorial paper Award (1995). He has been a visiting scholar at Queens University at Kingston, Canada (1991-1992) and at Cornell University, Ithaca, USA (2010-2011).

Thomas Chen is a Professor in Cyber Security at City, University of London. He received his BS and MS in electrical engineering from Massachusetts Institute of Technology, and PhD from University of California, Berkeley. He worked as senior member of staff at GTE Laboratories (now Verizon) in Waltham, Massachusetts. He joined Southern Methodist University in Dallas, Texas, as an associate professor, and then Swansea University, Wales, UK, as Professor in Networks. He has formerly served as editor-in-chief of IEEE Communications Magazine, IEEE Network, and IEEE Communications Surveys & Tutorials. He has co-written or co-edited seven books. He owns two US patents. His research interests are in network security, privacy, and cyber terrorism.